Deep Networks

Jin Sun

* Some figures are from Andrew Ng's cs294a course notes.
* Some figures are from Alex Krizhevsky's ImageNet paper.

Zoo of Networks

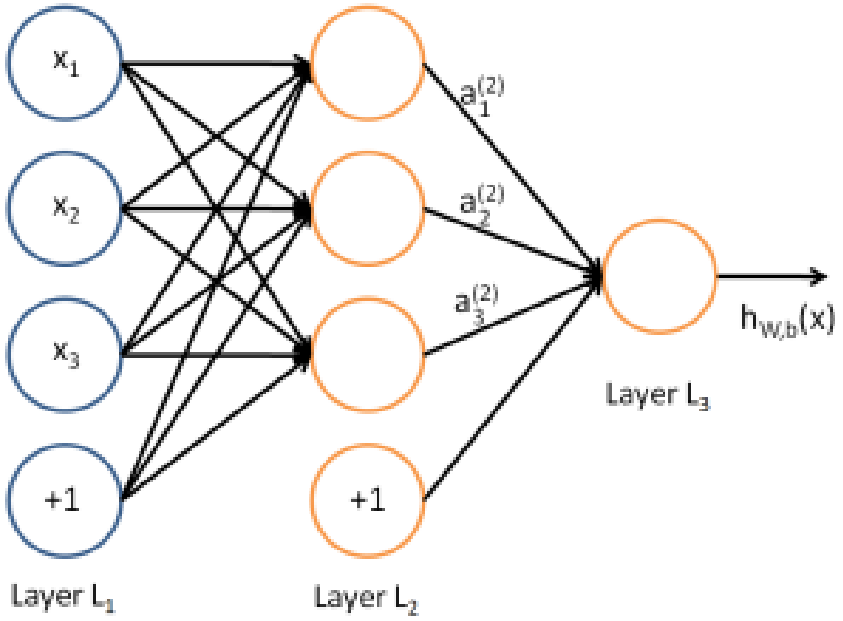
- Neural Networks
  - One layer neural network: logistic regression, perceptron
  - Multi-layer Perceptron, single hidden layer network: autoencoder
  - Deep neural networks
    - Convolutional Neural Networks: LeNet, ImageNet, R-CNN, MCDNN
    - Memory based neural networks: Recurrent Neural Networks, LSTM
- Belief Networks
  - Boltzmann Machine, Restricted Boltzmann Machine
  - Deep Boltzmann Machine
  - Deep Belief Networks

Key Concepts

- Backpropagation, SGD
- Non-linearity / Universal Approximator
- Non convexity / Generalization
- Going deep → exponentially powerful
- Vanishing (exploding) Gradient
- Content addressable memory
- Techniques
  - Generalization, constraints

Forward propagation

- From input layer to output layer
  - $a^{(l+1)} = f(W^{(l \to l+1)} a^{(l)})$
  - f is the activation (squashing) function
    - Non-linear: logistic, tanh, ReLU, etc…
    - Usually bounds intermediate values
    - Pros and Cons
- Loss function
  - Compute the error with target values
  - L2 loss, cross entropy loss



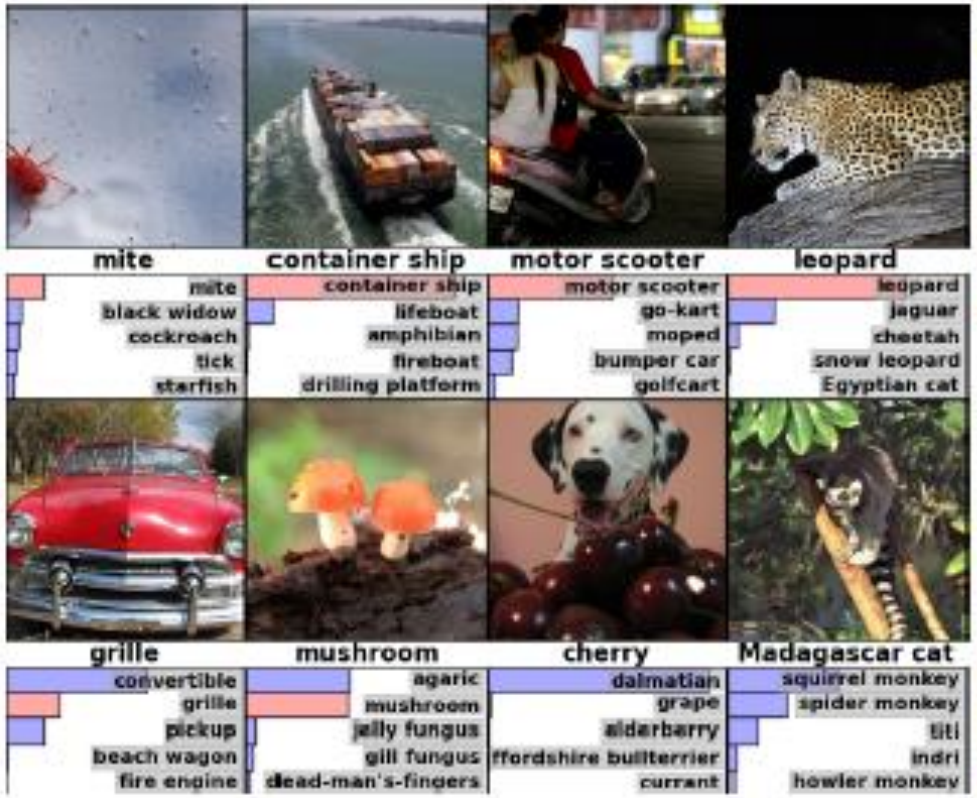
Backward propagation

- Gradient descent
  - $\frac{\partial loss}{\partial w^{(l \to l+1)}} = \left(\frac{\partial a^{(l+1)}}{\partial w^{(l \to l+1)}}\right)^T \left(\frac{\partial a^{(l+2)}}{\partial a^{(l+1)}}\right)^T \cdots \left(\frac{\partial y}{\partial a^{(last\ hidden\ layer)}}\right)^T \left(\frac{\partial loss}{\partial y}\right)$

- Momentum

$$\Delta w(t) = -lr * \nabla(w) + \eta \Delta w(t-1)$$

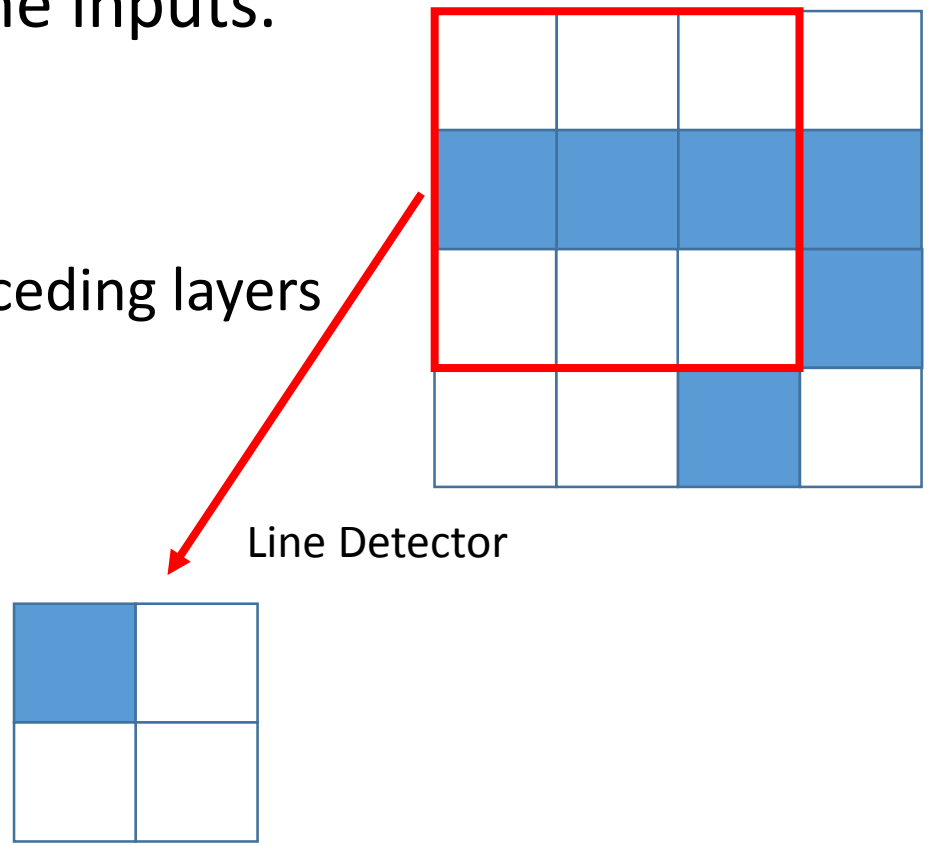
ImageNet



http://www.cs.toronto.edu/~fritz/absps/imagenet.pdf

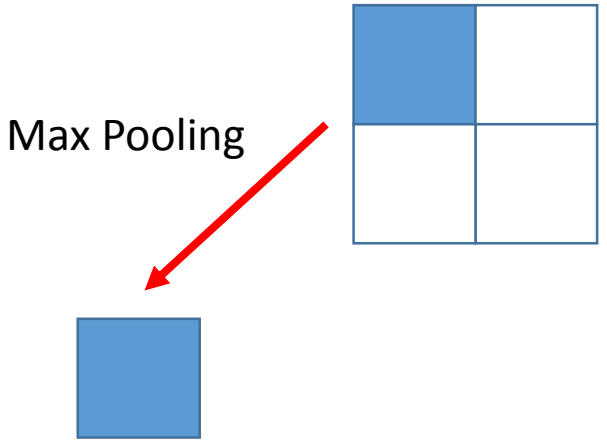
Convolutional Layer

- Fixed size filter (kernel) scanning on the inputs.
  - $a = f(\sum_{i,j \in kernel} w_{i,j} x_{r+i,c+j})$
- Detect local patterns
  - Low level features (lines, corners) in preceding layers
  - High level features (beard, hair, faces) in succeeding layers
- Network perspective:
  - sparse connection, identical weights



Pooling Layer

- Aggregating detections
  - The max (mean) value of the kernel
- Network perspective:
  - sparse connection, identical weights



Network Architecture

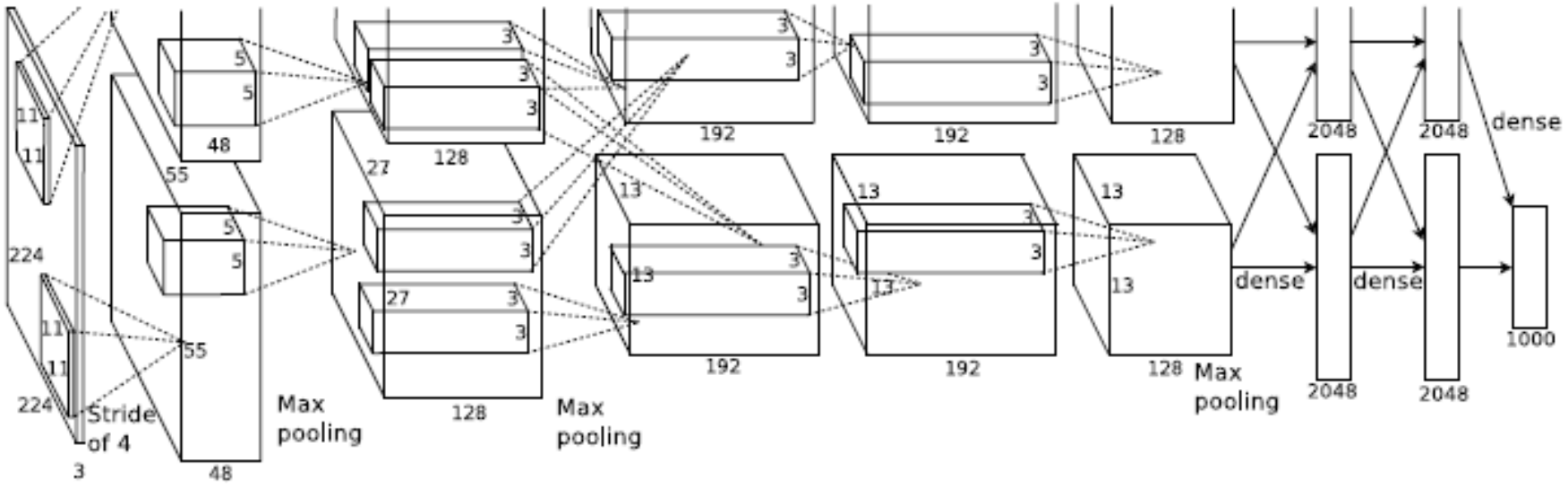
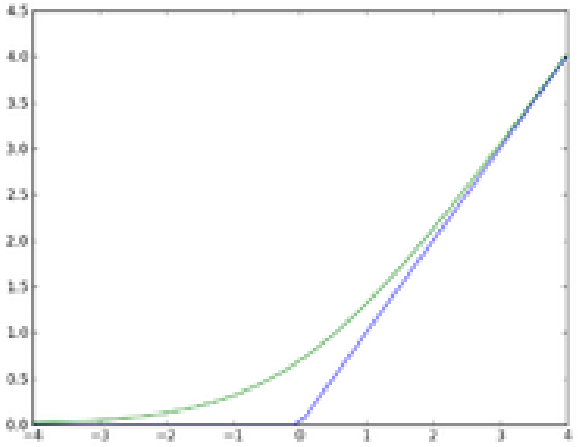
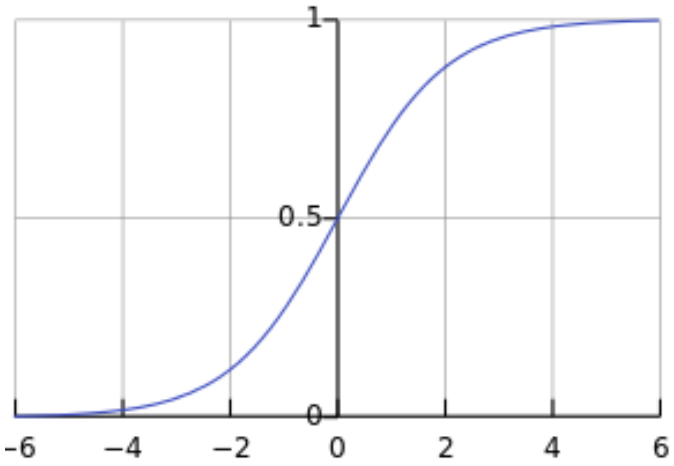


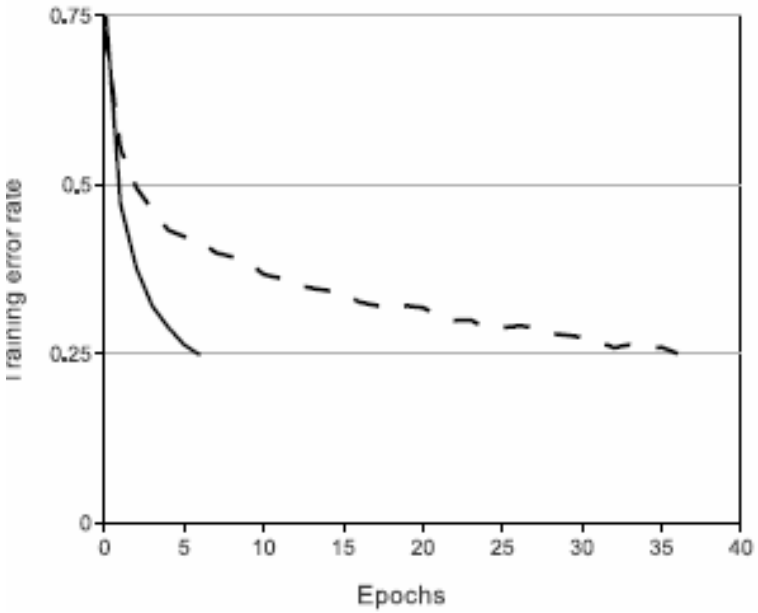
Figure 2: An illustration of the architecture of our CNN, explicitly showing the delineation of responsibilities between the two GPUs. One GPU runs the layer-parts at the top of the figure while the other runs the layer-parts at the bottom. The GPUs communicate only at certain layers. The network's input is 150,528-dimensional, and the number of neurons in the network's remaining layers is given by 253,440–186,624–64,896–64,896–43,264–4096–4096–1000.

ReLU activation

- $f(z) = \max(0, z)$
- Soft-plus: $f(z) = \ln(1 + e^z)$







Reduce Overfitting

- Data augmentation
  - Horizontal Reflection
  - PCA
    - Perform PCA on the set of RGB pixel values throughout the training set
    - For each RGB pixel, add $[p_1, p_2, p_3][\alpha_1\lambda_1, \alpha_2\lambda_2, \alpha_3\lambda_3]^{\mathsf{T}}$
    - $\alpha \sim \mathcal{N}(0, 0.1^2)$
    - Object identity is invariant to changes in the intensity and color of the illumination
- Dropout
- Local Response Normalization
  - $b_{x,y}^{i} = a_{x,y}^{i} / \left(k + \alpha \sum_{j=\max(0, i-n/2)}^{\min(N-1, i+n/2)} \left(a_{x,y}^{j}\right)^2\right)^{\beta}$
- Overlapping Pooling

Dataset

- ILSVRC: subset of ImageNet, ~1000 images in each of 1000 categories
  - 1.2 million training images
  - 50,000 validation images
  - 150,000 testing images
- Down-sample images to fixed resolution: 256*256
- Extract 224*224 patches from 256*256

Results

Model	Top-1	Top-5
Sparse coding [2]	*47.1%*	*28.2%*
SIFT + FVs [24]	*45.7%*	*25.7%*
CNN	**37.5%**	**17.0%**

Table 1: Comparison of results on ILSVRC-2010 test set. In *italics* are best results achieved by others.

Model	Top-1 (val)	Top-5 (val)	Top-5 (test)
SIFT + FVs [7]	—	—	*26.2%*
1 CNN	40.7%	18.2%	—
5 CNNs	38.1%	16.4%	**16.4%**
1 CNN*	39.0%	16.6%	—
7 CNNs*	36.7%	15.4%	**15.3%**

Table 2: Comparison of error rates on ILSVRC-2012 validation and test sets. In *italics* are best results achieved by others. Models with an asterisk* were "pre-trained" to classify the entire ImageNet 2011 Fall release. See Section 6 for details.